Enhanced Image Tampering Detection using Error Level Analysis and a CNN

Ramesh Gorle
Department of EECE, GITAM (Deemed to be University), Visakhapatnam, India
gorliramesh@gmail.com (corresponding author)

Anitha Guttavelli
Department of EECE, GITAM (Deemed to be University), Visakhapatnam, India
aguttave@gitam.edu (corresponding author)

Received: 12 November 2024 | Revised: 1 December 2024 | Accepted: 4 December 2024

 DOI: https://doi.org/10.48084/etasr.9593

ABSTRACT

This paper introduces a novel approach to image tampering detection by integrating Error Level Analysis (ELA) with a Convolutional Neural Network (CNN). Traditional forensic methods, such as ELA and Residual Pixel Analysis (RPA), often struggle to detect subtle or advanced manipulations in digital images. To address these limitations, this method leverages ELA to highlight compression-induced variations and CNN to extract and classify spatial features indicative of tampering. The dataset, consisting of both authentic and tampered images, was preprocessed to generate ELA representations, which were then used to train a CNN model designed to distinguish between authentic and manipulated regions. Extensive experimentation was performed on the CASIA v2.0 dataset, demonstrating significant improvements in detection accuracy, precision, and recall. The proposed framework achieved a detection accuracy of 96.21%, outperforming established deep learning models such as VGG16, VGG19, and ResNet101. These results underscore the potential of combining ELA and CNN in advancing image forensics, offering a robust solution to ensure the integrity of digital content in an era of sophisticated digital manipulation.

Keywords-error level analysis; convolution neural networks; image forensics; deep learning; digital image integrity

I. INTRODUCTION

In the digital age, maintaining the integrity and authenticity of visual content is increasingly important. With advances in image editing software, tampering has become widespread, creating challenges in ensuring information security and public trust. Manipulated images can mislead viewers, spread misinformation, and distort reality, leading to social and legal consequences. Therefore, detecting and identifying tampered areas within images is crucial to preserve the credibility of digital information and the trustworthiness of visual media.

Traditional methods of image forensics, such as Error Level Analysis (ELA) and Residual Pixel Analysis (RPA), have played a vital role in identifying tampering by examining artifacts introduced during editing or compression. ELA identifies discrepancies in compression levels across different parts of an image, while RPA analyzes pixel-level residuals. These techniques have provided forensic analysts with tools to detect tampering, helping to pinpoint altered areas. However, as manipulation methods become more advanced, traditional techniques face limitations, particularly in detecting subtle alterations and distinguishing between authentic and tampered regions. The evolving complexity of tamper methods highlights the need for more sophisticated and robust detection techniques to address increasingly intricate forgeries.

Recent advances in deep learning offer promising solutions to enhance image tampering detection. By extracting complex features from images, deep learning, specifically Convolutional Neural Networks (CNNs), can improve tampering detection. CNNs are well-suited for image forensics tasks due to their hierarchical learning capabilities. This study proposes a novel approach that integrates ELA with CNN technology, leveraging ELA's ability to emphasize error level variations and combining it with CNN's feature extraction capabilities to improve precision, accuracy, and reliability in tampering detection. Authentic and tampered images are preprocessed to generate ELA representations, capturing error-level variations introduced through editing or compression. These ELA images serve as input data for training a CNN model. This combination allows the model to learn to differentiate between authentic and tampered regions by recognizing distinct features within ELA images.

To create ELA images, the original image was saved in a lossy format (such as JPEG) with a specific quality setting, introducing compression artifacts that vary depending on content and alterations. Comparing the original image with its

compressed version reveals areas with differing compression levels, often indicative of tampering. These differences are scaled for visibility, allowing the CNN to learn relevant features.

The CNN architecture includes convolutional, pooling, and fully connected layers. Convolutional layers extract feature maps that highlight tampering patterns, while pooling layers reduce the spatial dimensions, making the model robust to input variations. A final classification layer outputs whether an image is tampered or authentic. This study divided the dataset into training and validation sets, applied data augmentation techniques such as flipping and rotating, and used binary cross-entropy as the loss function. The model was optimized with the Adam optimizer and early stopping to prevent overfitting, ensuring that it generalizes well to new data.

Image tampering deliberately alters images to deceive or manipulate viewers [1]. The increase in accessibility to powerful editing software has made image tampering a common practice. The consequences of image tampering can be severe, leading to misinformation, misrepresentation, and even legal issues [2]. Current methods for detecting image tampering rely mainly on deep learning techniques, ELA, and CNN. These methods have shown better accuracy than traditional approaches as they can extract complex features from images. Different techniques have been developed for this purpose, such as ELA and deep learning techniques. ELA detects tampering by comparing the compression level of one part of an image with another, figuring out the variations in the quality levels of different areas. The CNN is trained to figure out patterns for the detection of tampering.

Although these methods have shown promising results in detecting image tampering, they have certain limitations. One of the main challenges is the ability to detect copy-move forgery, where a part of an image is superimposed with another image. Another challenge is the ability to detect tampering in manipulated images that have undergone postprocessing to erase traces of tampering. Furthermore, detecting specific types of tampering, such as splicing, compression, rotation, or resampling, remains a challenge for many existing methods [3].

Deep learning techniques have played an important role in improving imatampering detectionion accuracy. These techniques can extract complex features from images, allowing more accurate tampering detection, and have shown better performance compared to traditional handcrafted feature-based methods. Additionally, deep learning methods can adapt well and generalize to different types of tampering operations, making them more robust in real-world scenarios [4]. In recent years, Large Language Models (LLMs) have gained prominence for their versatile applications in various domains. In [5], the role of LLMs in tamper detection was examined, focusing on their effectiveness in discerning AI-generated content and identifying image manipulations. By evaluating the capabilities of five major LLMs, this study shed light on the evolving landscape of tamper detection technology. Color Channel Characteristics (CCC) aid in detecting image tampering by analyzing noise and edge features across different color channels. The method extracts statistical features using various filters and edge detection techniques, capturing inconsistencies indicative of tampering. These features are classified using a Support Vector Machine (SVM), enabling accurate differentiation between authentic and manipulated images. The CCC approach enhances detection efficiency and accuracy in digital image forensics [6, 7].

In [8], a real-time tamper detection algorithm was presented that divided the image into parts and formed a pairwise correspondence. Other techniques for tamper detection and copyright protection in digital images focus on methods such as dual-redundant ring structures and fragile watermarking with block-neighborhood tamper detection. In addition, Non-Subsampled Contourlet Transform (NSCT) coefficients and Region of Interest (ROI) differentiation can be used for content restoration and tamper detection. In addition, this study emphasized the importance of data integrity and authenticity in medical image processing along with the importance of watermarking to ensure the security of sensitive medical data.

Image quality is one of the important features in tamper detection. In [9], various digital image watermarking techniques were reviewed, emphasizing their importance in protecting digital content from unauthorized access and manipulation. This study also discussed classification based on working domains and applications, highlighted challenges in achieving robustness and imperceptibility, and outlined future research directions in the field of digital watermarking. In [10], a CNN-based method was proposed for real-time image quality enhancement. In [11], a DenseNet model was proposed to address the limitations of typical CNNs in detecting forged images, achieving an accuracy of 92.32%. In [12], ELA and Scale-Invariant Feature Transform (SIFT) features were combined to enhance image authentication by comprehensively analyzing error patterns and distinctive key points. This approach offered robustness against various image transformations, precise localization of manipulations, improved classification accuracy, and practical implementation potential for real-world applications.

In [13], a two-channel progressive feature filtering network was proposed to detect tampered images and localize the manipulation. In [14], ELA was used with CNN to detect image forgeries, based on the compression ratio between the fake and original images. Similarly, in [15], ELA was used with a CNN to detect tampered images. In [16], lightweight CNN was used to detect splicing forged images in real time. In [17], a two-phase approach was proposed, using an ensemble of VGG-16 and CNN architectures to detect image forgeries. This approach combines ELA with deep learning methods to improve accuracy in distinguishing between original and manipulated images. This study used a Kaggle dataset, preprocessed images, extracted features using CNN, and trained the model. The test results showed high accuracy in detecting artificial distortions, demonstrating the effectiveness of the proposed method in image tampering detection.

II. PROPOSED FRAMEWORK

The proposed method aims to accurately identify tampered images from authentic ones. This method is divided into several key phases, as shown in Figure 1.

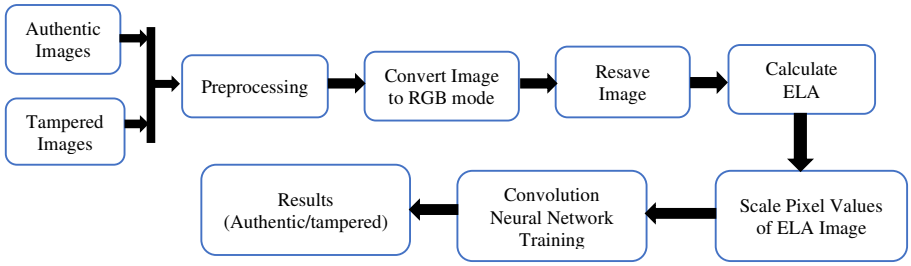

Fig. 1. The proposed method.

The CASIA 2.0 dataset [18] was used, which consists of authentic and tampered images. This dataset includes 12,614 images in four formats, BMP, JPG, PNG, and TIF, where 7,491 are original photos and 5,123 are tampered. CASIA 2.0 represents various image types, such as animals, buildings, articles, characters, plants, nature, sceneries, textures, and interior shots. This dataset contains images in varying sizes and resolutions, from 800×600 to 384×256 pixels. Table I shows the distribution of images in the dataset. Figures 2 and 3 show some of the authentic and tampered images from the dataset.

TABLE I. DETAILS OF THE DATASET

Dataset	Image formats	Authentic images	Tampered images
CASIA v2.0	jpg, tif, bmp	7492	5123

Fig. 2. Authentic images from the dataset.

Fig. 3. Tampered images from the dataset.

A. Data Preprocessing

Data preprocessing is one of the most important parts, preparing images for a deep learning model's training. This study performed the following preprocessing steps:

- Resizing: All images were resized to the same size 128×128 pixels so that the input dimensions are the same for all.
- Normalization: Pixel intensities were normalized between [0, 1] to work on a more balanced representation while training.
- Data Augmentation: Artificially increase the size of the training dataset by applying techniques such as rotation, flipping, and zooming.
- Standardization: Converting all images to RGB ensures that they have consistent color spaces. Different images might come in different formats (e.g., grayscale), and standardizing them to RGB makes the subsequent processing steps simpler and more uniform.

B. Error Level Analysis (ELA)

ELA is an image forensic technique that can be utilized to highlight the differences between the original image and its compressed version to detect manipulations. This is based on the fact that different parts of the image compress differently depending on their content. ELA allows one to identify such differences to detect tampered regions.

The ELA algorithm starts by letting $I_{orig}(x, y)$ represent the pixel intensity of the original image at (x, y) coordinates. The original image is then re-saved at a known compression level, resulting in a compressed image. $I_{comp}(x, y)$ represents the pixel intensity of this recompressed image at the same coordinates (x, y). ELA is performed by calculating the absolute difference between the original image and its recompressed version. This difference highlights areas with differing compression levels, possibly being the tampered regions.

$$I_{ELA}(x, y) = \left| I_{orig}(x, y) - I_{comp}(x, y) \right| \quad (1)$$

where $I_{ELA}(x, y)$ represents the pixel intensity of the ELA image at the coordinates (x, y). The resulting ELA image is scaled to enhance visibility. The maximum difference in the ELA image determines the scaling factor.

$$scale = \frac{255}{max(I_{ELA}(x,y))} \quad (2)$$

The scaled ELA image is then enhanced by

$$I_{ELA}(x, y) = I_{ELA}(x, y)\ X\ scale \quad (3)$$

Figure 4 shows examples of authentic and tampered images and their respective ELA images.

C. CNN Model Architecture

A CNN was used to detect tampering. This model is a sequential stack of layers where each has exactly one input tensor and one output tensor. It consists of several layers, as shown in Figure 5. The model starts with a Conv2D layer applying 32 filters with a 5×5 kernel, 'valid' padding, and ReLU activation to input images of shape (128, 128, 3), followed by another Conv2D layer with the same parameters.

Then, it includes a MaxPool2D layer with a 2×2 pool size to reduce spatial dimensions and a rate of dropout layer with a 0.25 rate to prevent overfitting. The output is flattened and passed through a dense layer with 256 units and ReLU activation, followed by another dropout layer with a 0.5 rate. Finally, it has a dense output layer with 2 units and softmax activation for binary classification. The model uses the training dataset to train it, and its performance is validated on a separate validation dataset.

The whole process includes the following steps:

- Splitting the data: The dataset is divided into training and validation sets in the ratio 80:20.
- Model compilation: This model is compiled with the Adam optimizer with a learning rate of 0.001. The loss used is the binary cross-entropy loss.
- Training: The model is trained for 30 epochs using early stopping as the mechanism to avoid overfitting.

Evaluation: Accuracy, precision, recall, and F1-score metrics evaluate the model's performance. In addition, a confusion matrix is used to illustrate the classification performance.

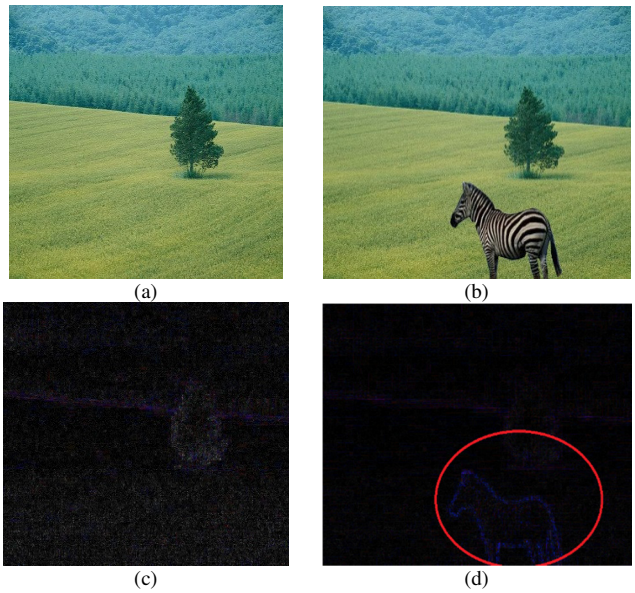

Fig. 4. (a) Authentic image, (b) Tampered image, (c) ELA of the authentic image, (d) ELA of the tampered image.

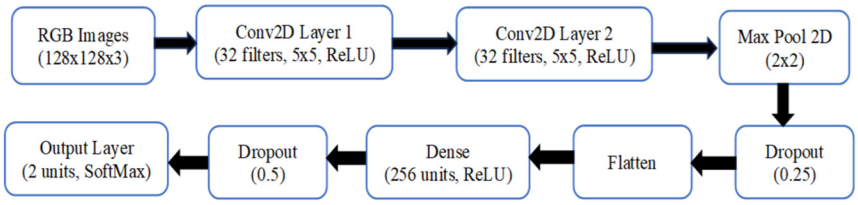

Fig. 5. The proposed CNN model.

Figure 6 shows the training and validation loss and accuracy over 20 epochs of the proposed model. Training and validation losses decrease over time, indicating that the model is learning well.

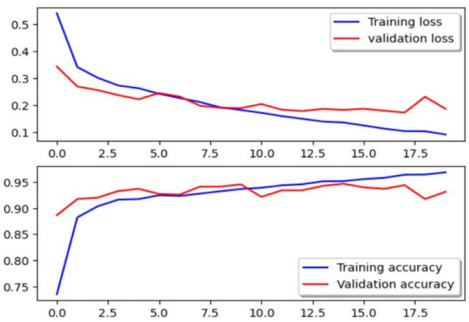

Fig. 6. Training loss and accuracy curves of the proposed model.

III. RESULTS AND ANALYSIS

The performance analysis was based on accuracy, precision, recall, and F1 score. The following terms are used to calculate these parameters.

- TP (True Positive): The amount of tampered images that are correctly identified as tampered.
- TN (True Negative): The amount of authentic images that are correctly identified as authentic.
- FN (False Negative): The amount of tampered images that are incorrectly identified as authentic.
- FP (False Positive): The amount of authentic images that are incorrectly identified as tampered.

Accuracy is the ratio of correctly classified images to total images.

$$\text{Accuracy} = \frac{(TP+TN)}{(TP+TN+FP+FN)} \quad (4)$$

Precision is the ratio of TP to the total number of images classified as positive.

$$\text{Precision} = \frac{TP}{(TP+FP)} \quad (5)$$

Recall is the ratio of TP images to the total number of positive images.

$$\text{Recall} = \frac{TP}{(TP+FN)} \quad (6)$$

F1 score is the harmonic mean of precision and recall.

$$\text{F1 Score} = 2\,\frac{\text{Precision}\times\text{Recall}}{\text{Precision}+\text{Recall}} \quad (7)$$

Figure 7 shows a bar chart that compares the accuracy of four different models in detecting image tampering. The models evaluated are the proposed CNN, VGG16, VGG19, and ResNet101. The proposed model achieved the highest accuracy at 96.21%, significantly outperforming the others. VGG19 and VGG16 also showed strong performance with 90.32% and 88.92% accuracy, respectively. However, ResNet101 lagged with an accuracy of only 74.75%. This comparison highlights the superior effectiveness of the proposed CNN model in accurately identifying tampered images.

Fig. 7. Accuracy comparison.

Figure 8 illustrates the precision of the four models. The proposed model achieved the highest precision of 98.58%, indicating its high reliability in identifying tampered images. VGG16 followed with a precision of 92.69%, while VGG19 recorded a precision of 88.16%. The precision metric emphasizes the exceptional performance of the proposed model in minimizing FP.

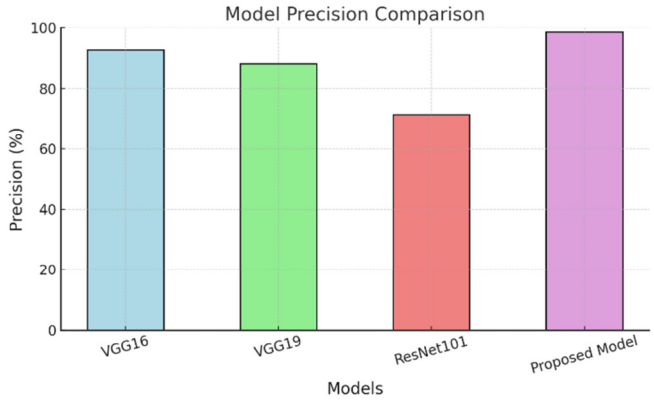

Fig. 8. Precision comparison.

Figure 9 shows a recall comparison bar chart, indicating how effectively the models identified actual tampered images from the dataset. The proposed model led with a recall rate of 92.36%, indicating its ability to detect most of the tampered images. VGG19, VGG16, and ResNet101 had lower recall rates of 89.04%, 80.07%, and 66.78%, respectively. The high recall of the proposed model signifies its robustness in identifying tampered images.

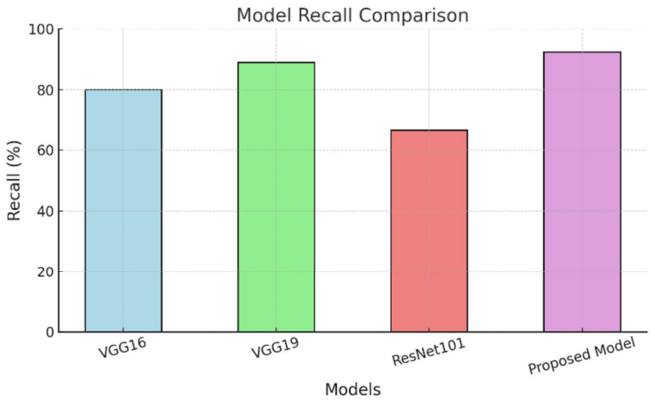

Fig. 9. Recall comparison.

Figure 10 compares the F1 score of four different models in detecting tampered images. The proposed model achieved the highest F1 score of 95.37%, significantly outperforming the others.

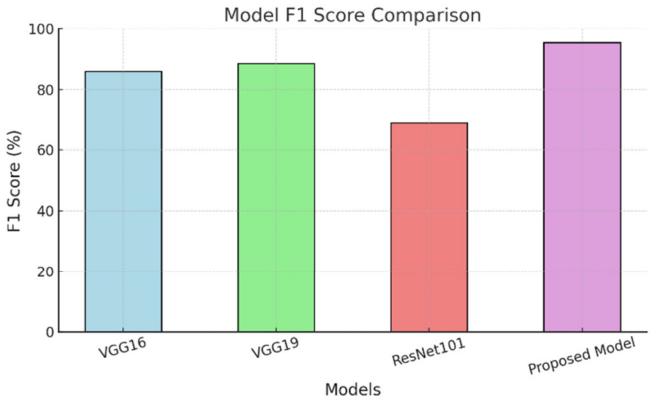

Fig. 10. F1-score comparison.

VGG19 and VGG16 also showed strong performance with F1 scores of 88.60% and 85.92%, respectively. ResNet101 lagged with an F1 score of only 69.07%. This comparison highlights the superior effectiveness of the proposed CNN model in accurately identifying tampered images. Table II presents a detailed performance analysis of the four models.

TABLE II. PERFORMANCE COMPARISON OF THE FOUR MODELS

Model	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
VGG16	88.92	92.69	80.07	85.92
VGG19	90.32	88.16	89.04	88.60
ResNet101	74.75	71.53	66.78	69.07
Proposed Model	96.21	98.58	92.36	95.37

Using CNN along with ELA, the confidence level of one real and one fake image was improved to 99.70% and 99.89%, respectively, as shown in Figures 11 and 12. Figure 13 shows the confusion matrix, detailing the performance of the proposed CNN. The recall was also substantial. However, there was a moderate number of FN (23), indicating that some original samples were misclassified as tampered. Overall, the model demonstrated strong performance in identifying both fake and real samples.

Fig. 11. Prediction and confidence of a real image.

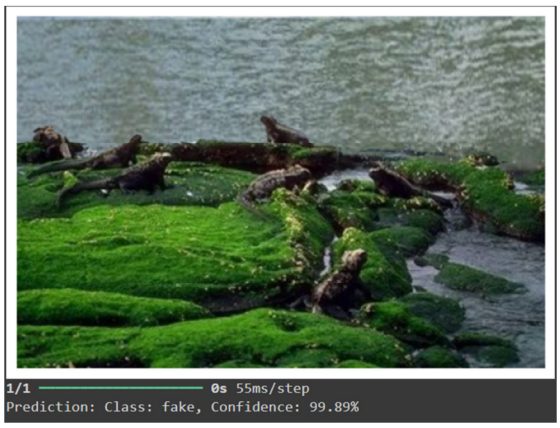

Fig. 12. Prediction and confidence of a fake image.

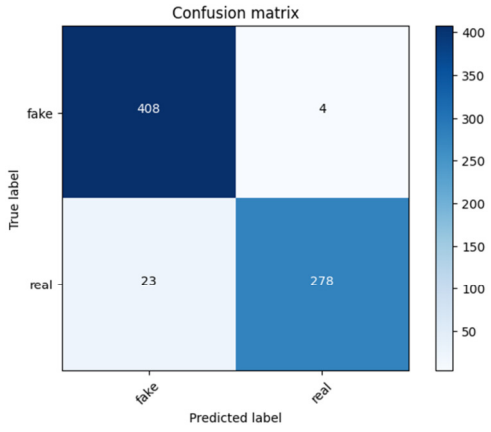

Fig. 13. Confusion matrix.

IV. CONCLUSION

This study introduced a novel framework for detecting tampered images by combining the strengths of ELA and CNNs. The proposed approach stands out for its ability to address the limitations of traditional forensic methods, such as Residual Pixel Analysis (RPA) and standalone ELA, which struggle to detect subtle or sophisticated manipulations. By leveraging ELA's capability to highlight compression-induced variations and CNN's prowess in extracting high-level spatial features, the proposed method achieved a robust solution for tampering detection.

This research contributes significantly to the field of image forensics by demonstrating a marked improvement in performance metrics, including accuracy, precision, recall, and F1 score. Experimental results on the CASIA v2.0 dataset highlight the superiority of the proposed framework, achieving 96.21% accuracy, 98.58% precision, 92.36% recall, and 95.37% F1 score, surpassing established deep learning models such as VGG16, VGG19, and ResNet101. This advance can be attributed to the effective preprocessing of images through ELA and the utilization of a carefully designed CNN architecture capable of identifying complex patterns indicative of tampering. This work not only advances the state-of-the-art in image tampering detection but also sets a foundation for future research to explore similar hybrid approaches in other areas of forensic analysis.

DECLARATIONS

A. *Conflict of Interest*

The authors declare that this research does not cause any conflict of interest.

B. *Funding*

The authors did not receive financial support or funding that could influence the findings or conclusions of this research.

REFERENCES

[1] G. Li, X. Yang, and W. Ma, "A Two-Stage Dual-Path Framework for Text Tampering Detection and Recognition." arXiv, Feb 22, 2024, https://doi.org/10.48550/arXiv.2402.13545.

[2] R. Gorle and A. Guttavelli, "A novel dynamic image watermarking technique with features inspired by quantum computing principles," *AIP Advances*, vol. 14, no. 4, Apr. 2024, Art. no. 045024, https://doi.org/10.1063/5.0209417.

[3] J. Xing, X. Tian, and Y. Han, "A Dual-channel Augmented Attentive Dense-convolutional Network for power image splicing tamper detection," *Neural Computing and Applications*, vol. 36, no. 15, pp. 8301–8316, May 2024, https://doi.org/10.1007/s00521-024-09511-6.

[4] S. Chopparapu and J. B. Seventline, "An Efficient Multi-modal Facial Gesture-based Ensemble Classification and Reaction to Sound Framework for Large Video Sequences," *Engineering, Technology & Applied Science Research*, vol. 13, no. 4, pp. 11263–11270, Aug. 2023, https://doi.org/10.48084/etasr.6087.

[5] X. Yang and J. Zhou, "Research about the Ability of LLM in the Tamper-Detection Area." arXiv, Jan. 24, 2024, https://doi.org/10.48550/arXiv.2401.13504.

[6] S. Gupta and N. Mohan, "Color Channel Characteristics (CCC) for Efficient Digital Image Forensics," *Engineering, Technology & Applied Science Research*, vol. 8, no. 1, pp. 2555–2561, Feb. 2018, https://doi.org/10.48084/etasr.1744.

[7] C. SaiTeja and J. B. Seventline, "A hybrid learning framework for multi-modal facial prediction and recognition using improvised non-linear SVM classifier," *AIP Advances*, vol. 13, no. 2, Feb. 2023, Art. no. 025316, https://doi.org/10.1063/5.0136623.

[8] V. Sisaudia and V. P. Vishwakarma, "Approximate regeneration of image using fragile watermarking for tamper detection and recovery in real time," *Multimedia Tools and Applications*, vol. 83, no. 25, pp. 66299–66318, Jul. 2024, https://doi.org/10.1007/s11042-024-18247-3.

[9] M. Begum and M. S. Uddin, "Digital Image Watermarking Techniques: A Review," *Information*, vol. 11, no. 2, Feb. 2020, Art. no. 110, https://doi.org/10.3390/info11020110.

[10] S. Chopparapu, G. Chopparapu, and D. Vasagiri, "Enhancing Visual Perception in Real-Time: A Deep Reinforcement Learning Approach to Image Quality Improvement," *Engineering, Technology & Applied Science Research*, vol. 14, no. 3, pp. 14725–14731, Jun. 2024, https://doi.org/10.48084/etasr.7500.

[11] A. Alzahrani, "Digital Image Forensics: An Improved DenseNet Architecture for Forged Image Detection," *Engineering, Technology & Applied Science Research*, vol. 14, no. 2, pp. 13671–13680, Apr. 2024, https://doi.org/10.48084/etasr.7029.

[12] R. M. Bora and M. R. Sanghavi, "Feature Extraction and Analysis of SIFT features for ELA of Authentic and Forged Images," *International Journal of Intelligent Systems and Applications in Engineering*, vol. 12, no. 16s, pp. 368–373, Feb. 2024.

[13] S. Fu, Q. Chen, Z. LI, G. Wang, and Y. Lu, "Two-channel progressive feature filtering network for tampered image detection and localization," *Journal of Computer Applications*, vol. 44, no. 4, 2024, Art. no. 1303.

[14] D. J. Kotti, D. E. Gouthami, D. K. Swapna, and S. Vesalapu, "Morphed Image Detection Using ELA and CNN Techniques," *Journal of Pharmaceutical Negative Results*, pp. 4069–4078, Dec. 2022.

[15] I. B. K. Sudiatmika, F. Rahman, T. Trisno, and S. Suyoto, "Image forgery detection using error level analysis and deep learning," *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, vol. 17, no. 2, pp. 653–659, Apr. 2019, https://doi.org/10.12928/telkomnika.v17i2.8976.

[16] K. M. Hosny, A. M. Mortda, N. A. Lashin, and M. M. Fouda, "A New Method to Detect Splicing Image Forgery Using Convolutional Neural Network," *Applied Sciences*, vol. 13, no. 3, Jan. 2023, Art. no. 1272, https://doi.org/10.3390/app13031272.

[17] S. B. N, G. T. M. Kumar, M. H. C, A. M, and N. Sneha, "Image Forgery Detection Using Ensemble of VGG-16 and CNN Architecture," *International Journal of Computational Learning & Intelligence*, vol. 2, no. 3, pp. 122–135, Aug. 2023, https://doi.org/10.5281/zenodo.8210388.

[18] D. Goel, "CASIA 2.0 Image Tampering Detection Dataset." Kaggle, [Online]. Available: https://www.kaggle.com/datasets/divg07/casia-20-image-tampering-detection-dataset.